

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## A combinatorial approach to angiosperm pollen morphology

### Journal Item

#### How to cite:

Mander, Luke (2016). A combinatorial approach to angiosperm pollen morphology. Proceedings of the Royal Society B: Biological Sciences, 283(1843), article no. 20162033.

For guidance on citations see [FAQs](#).

© 2016 The Author



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.1098/rspb.2016.2033>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

1    **A combinatorial approach to angiosperm pollen morphology**

2

3    Luke Mander<sup>1</sup>

4

5    <sup>1</sup>Department of Environment, Earth and Ecosystems, The Open University,

6    Milton Keynes, MK7 6AA, UK

7

8    **Author for correspondence:**

9    Luke Mander

10    e-mail: luke.mander@open.ac.uk

11

12    **Abstract**

13    Angiosperms (flowering plants) are strikingly diverse. This is clearly expressed

14    in the morphology of their pollen grains, which are characterised by enormous

15    variety in their shape and patterning. In this paper, I approach angiosperm

16    pollen morphology from the perspective of enumerative combinatorics. This

17    involves generating angiosperm pollen morphotypes by algorithmically

18    combining character states and enumerating the results of these combinations. I

19    use this approach to generate 3,643,200 pollen morphotypes, which I visualise

20    using a parallel coordinates plot. This represents a raw morphospace. In order to

21    compare real-world and theoretical morphologies, I map the pollen of 1008

22    species of Neotropical angiosperms growing on Barro Colorado Island (BCI),

23    Panama, onto this raw morphospace. This highlights that in addition to their

24    well-documented taxonomic diversity, Neotropical rainforests also represent an

25    enormous reservoir of morphological diversity. Angiosperm pollen morphospace

26 at BCI has been filled mostly by pollen morphotypes that are unique to single  
27 plant species. Repetition of pollen morphotypes among higher taxa at BCI  
28 reflects both constraint and convergence. This combinatorial approach to  
29 morphology addresses the complexity that results from large numbers of  
30 discrete character combinations, and could be employed in any situation where  
31 organismal form can be captured by discrete morphological characters.

32

33 **Keywords:**

34 plants, morphology, morphospace, combinatorics, pollen

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51 **1. Introduction**

52 Angiosperms (flowering plants) are an extremely diverse group of terrestrial  
 53 plants and are composed of an estimated 260,000 species [1]. Angiosperms are  
 54 characterised by striking morphological diversity, and the group contains an  
 55 array of life forms that includes herbs, epiphytes, bulbs, aquatic plants, shrubs  
 56 and trees. Research into the morphological diversity of angiosperms has  
 57 involved the development of morphospaces that allow the morphology of  
 58 different taxa to be quantitatively compared [2]. For example, morphospaces for  
 59 angiosperm flowers have been used to investigate the frequency with which  
 60 certain flower morphologies occur in nature [3] and also to examine the  
 61 disparity of floral shapes [4]. Studies of fossil angiosperm pollen grains have  
 62 shown that the taxonomic diversity and morphological disparity of angiosperms  
 63 were decoupled during the evolutionary radiation of this clade during the  
 64 Cretaceous period (~150–65 million years ago) [5], and recent work has  
 65 highlighted that plant clades, including the angiosperms, typically attain high  
 66 level of morphological disparity early in their evolutionary history [2].

67 In general, researchers recognise two approaches to the construction of a  
 68 morphospace. The first approach is theoretical, which involves using a  
 69 mathematical model to generate morphologies and may also include comparison  
 70 with real-world forms [6,7]. In this context, simulations of plant phenotypes [8]  
 71 are a botanical counterpoint to simulations of animal shell shape [9], and are a  
 72 classic example of theoretical morphospace construction. The second approach  
 73 is empirical, which involves constructing a morphospace from a set of empirical  
 74 morphological data [2]. In an empirical morphospace, the form of an organism is  
 75 typically represented by a number of discrete characters, and there is often

76 overlap between these characters and the set of characters that could be used for  
77 the purpose of classification or phylogenetic analysis [10]. The dimensionality of  
78 an empirical morphospace is usually reduced using multivariate ordination [6,7],  
79 and the morphological disparity of a clade can be tracked through time using a  
80 distance metric such as mean pairwise dissimilarity [e.g. 5,10]. The  
81 morphospaces for plants constructed by Lupia [5] and Oyston et al. [2] are  
82 empirical, and in this respect they are similar to foundational studies on the  
83 morphological diversity of marine animals that are also based on discrete  
84 characters [e.g. 10,11].

85       In addition to these theoretical and empirical approaches, however, there  
86 are also raw morphospaces, which are formulated in terms of observed  
87 morphological variation but prior to any multivariate ordination [7]. These  
88 morphospaces are empirical in the sense that aspects of real form are captured  
89 by the enumeration of discrete variables, rather than using a generative model,  
90 but they are also theoretical in the sense that they can represent forms that do  
91 not exist in nature [7,12]. Raw morphospaces are useful tools in the investigation  
92 of organic form because they facilitate "the study of the evolution of the actual in  
93 the realm of the possible" [7, p. 10]. Stebbins produced a raw morphospace for  
94 angiosperm flowers by constructing a grid measuring 16 cells by 16 cells [3].  
95 Each of the 256 cells in this grid represented one of 256 possible combinations of  
96 primitive versus advanced conditions for the eight binary floral characters that  
97 Stebbins analysed [3]. This grid allowed Stebbins to graphically represent  
98 morphological types produced by different discrete character combinations, and  
99 to demonstrate that the proportion of floral shapes that exist in nature is a  
100 limited subset of the floral shapes that might be structurally possible [3]. For

101 Stebbins, one of the primary problems with his analysis was “the distortion  
102 produced by the recognition of only two conditions in respect to each character.  
103 This [he noted] was necessary because of the great complexity which would have  
104 resulted if more conditions had been recognized” [3, p. 302].

105       In this paper I address the complexity that results from large numbers of  
106 discrete character combinations by approaching angiosperm pollen morphology  
107 from the perspective of enumerative combinatorics. I begin with a schematic  
108 example that demonstrates the construction of a raw morphospace for  
109 angiosperm pollen. I use this approach to study angiosperm pollen morphology  
110 in the hyperdiverse Neotropics, and this involves comparing the morphology of  
111 1,008 real-world pollen grains to 3,643,200 algorithmically generated pollen  
112 morphotypes. The approach I outline is not restricted to angiosperm pollen, and  
113 could be employed in any situation where organismal form can be captured by  
114 discrete morphological characters.

115

## 116 **2. A combinatorial approach to angiosperm pollen morphology**

117 The overall morphology of a single angiosperm pollen grain is determined by the  
118 nature and arrangement of several different individual characters. The states of  
119 these characters are typically described using specialist terminology, which  
120 allows researchers to communicate information about pollen [e.g. 13]. This  
121 specialist terminology treats each of these character states as discrete elements  
122 that contribute to the overall morphology of a pollen grain. Different pollen  
123 morphotypes reflect different combinations of these character states. This is  
124 shown in Figure 1a that contains five morphological characters, each with a

125 number of different states, which can be used to construct a schematic pollen  
126 grain (descriptive terminology in this paper follows [13]).

127

## 128 **2.1. Counting the number of possible angiosperm pollen morphotypes**

129 One possible combination of the character states shown in Figure 1a is: 3, 2, 2, 1,  
130 2, 4. In the language of formal descriptive palynology [13], this combination  
131 would produce a tricolpate pollen grain with a tectate exine, spheroidal shape,  
132 monad dispersal, and an echinate–reticulate surface ornamentation. In more  
133 general terminology, this pollen grain has three elongated apertures in its exine,  
134 which is the outer coating of a pollen grain that is composed of sporopollenin  
135 [13] and in this case possesses a distinctive layer known the tectum that forms a  
136 roof over its internal structures [13]. This pollen grain has a spherical 3-  
137 dimensional shape, is dispersed at maturity by its parent plant as a single grain  
138 (a monad), and has spines longer than 1µm together with a network-like pattern  
139 on the surface of its exine. Generating this combination involves selecting one  
140 state from each list of aperture, exine, shape and dispersal character states. Some  
141 pollen grains possess just a single type of surface ornamentation, but others have  
142 two distinct types that form primary and secondary surface ornamentation  
143 (although the two are not hierarchically related). *Razisea spicata* (Acanthaceae),  
144 for example, has sparse echini and dense baculae [see 13 for terminology].  
145 Accordingly, this combination involves selecting two states from the list of  
146 surface ornamentation character states.

147 I have thought of this process of character state selection as a  
148 combinatorial counting problem, in which the number of ways of selecting  $k$   
149 objects from a list of  $n$  objects, if the order of selection is irrelevant, is given by

$$\binom{n}{k} = \begin{cases} \frac{n!}{k!(n-k)!} & \text{if } 0 \leq k \leq n, \\ 0 & \text{otherwise.} \end{cases}$$

150 The numbers  $\binom{n}{k}$ , which can be read " $n$  choose  $k$ ", are the binomial coefficients,  
 151 and the value of  $\binom{n}{k}$  is 0 when  $k < 0$  or  $k > n$  because there are no ways to select  
 152 fewer than zero or more than  $n$  objects from a list of  $n$  objects [14]. In the  
 153 schematic example shown in Figure 1, the values of  $\binom{n}{k}$  for each of the five  
 154 characters are shown in Figure 1b.

155 For the aperture, exine, shape and dispersal characters, only one state is  
 156 selected from each character state list, and the binomial coefficients for these  
 157 characters are written  $\binom{n}{1}$  (Fig. 1b). For the aperture character, for example,  
 158 there are  $\binom{3}{1} = 3$  ways to select a single character state from this list. For surface  
 159 ornamentation, two states are chosen from the list of five character states, and  
 160 the binomial coefficients for this character are written  $\binom{5}{2}$  (Fig 1b). In this case  
 161 there are  $\binom{5}{2} = 10$  ways to select two states from this list. By the product rule,  
 162 the number of all possible combinations of these character states is

$$\binom{3}{1} \binom{2}{1} \binom{3}{1} \binom{2}{1} \binom{5}{2} = 360.$$

163 In other words, there are 360 unique pollen morphotypes that can be generated  
 164 from the five characters shown in Figure 1a using  $n$  choose  $k$  combination.  
 165

## 166 **2.2. Enumerating and displaying angiosperm pollen morphotypes**

167 For small values of  $\binom{n}{k}$  it is possible to count and enumerate combinations  
 168 manually. For example, there are six combinations of two letters that can be  
 169 produced from the four letters ABCD, and this is given by  $\binom{4}{2}$  or 4 choose 2. This



170 can be calculated using factorials or by using Pascal's triangle for binomial  
171 coefficients, and these six combinations can be enumerated manually by hand:  
172 AB AC AD BC BD CD. However, manual counting and enumerating becomes  
173 rapidly intractable as the values of  $\binom{n}{k}$  increase. In order to overcome this, I have  
174 written an algorithm using the Python programming language that generates  
175 angiosperm pollen morphotypes by  $n$  choose  $k$  combination of morphological  
176 characters. An example of this algorithm, which counts and enumerates all  
177 possible combinations of the character states shown in Figure 1a, is provided in  
178 the Supplementary Material. This algorithm takes lists of morphological  
179 characters as it's input, enumerates the combinations of these morphological  
180 characters using  $n$  choose  $k$  combination, and then writes these combinations to  
181 a .csv file with headers for each character. An example of a .csv file containing all  
182 possible combinations of the character states shown in Figure 1a is provided in  
183 the Supplementary Material.

184 I have visualised these character combinations using a parallel  
185 coordinates plot (Fig. 1c). In this plot, each vertical axis represents a different  
186 morphological character and lines visualise associations between character  
187 states. Parallel coordinates plots can be thought of as multipartite graphs, with  
188 each vertical axis corresponding to a graph partition [15], and they have been  
189 used to visualise large volumes of network traffic data [16]. This plot represents  
190 a raw morphospace that is bounded and contains a range of possible  
191 morphologies (Fig 1c). Provided that the same characters and states are used,  
192 real-world pollen grains can be mapped onto this morphospace.

193

### 194 3. Hyperdiversity in the Neotropics

195 I have applied this combinatorial approach to the hyperdiverse angiosperm-  
 196 dominated tropical rainforests of South America. This region of the world  
 197 contains the highest levels of plant diversity on Earth, and supports an estimated  
 198 ~16,000 tree species alone [17]. When both extant and extinct taxa are  
 199 considered, the total diversity of all Neotropical plant species that have ever  
 200 existed is potentially enormous. I have focussed on the morphology of  
 201 angiosperm pollen grains produced by plants growing on Barro Colorado Island  
 202 (BCI). Barro Colorado Island is a 1,560-ha island situated in the Panama Canal  
 203 that supports hyperdiverse lowland moist tropical forest, and a 50-ha plot on BCI  
 204 that has been monitored since 1980 contains just over 300 tree and shrub  
 205 species [18]. This region is of particular interest in the context of this present  
 206 paper because such high species diversity offers an opportunity to confront the  
 207 challenge of analysing large numbers of discrete character combinations [3]. The  
 208 vegetation of BCI also represents a single well-studied flora with which to  
 209 examine the patterns of morphospace occupation by different plant groups, and  
 210 to explore the properties of occupied and unoccupied morphospace [e.g. 12].

211

### 212 **3.1. Morphospace construction**

213 I used the methods outlined in Figure 1 to construct a raw angiosperm pollen  
 214 morphospace based on 60 aperture character states, four exine character states,  
 215 ten shape character states, six dispersal character states and 23 surface  
 216 ornamentation character states (a full list of character states is provided in the  
 217 Supplementary Material). For surface ornamentation, two states are chosen from  
 218 the list of 23 character states, and there are  $\binom{23}{2} = 253$  ways to select two states

219 from this list. By the product rule, the number of all possible combinations of  
 220 these character states is

$$\binom{60}{1} \binom{4}{1} \binom{10}{1} \binom{6}{1} \binom{23}{2} = 3,643,200.$$

221 The raw morphospace that represents these 3,643,200 algorithmically generated  
 222 pollen morphotypes is shown in Figure 2, and this allows associations between  
 223 character states to be visualised. These character states were selected from the  
 224 glossary of Punt et al. [13], and each is known to exist in nature. Some  
 225 combinations of these character states are also known to exist in nature. For  
 226 example, the combination Aperture 3, Exine 4, Shape 9, Dispersal 1,  
 227 Ornamentation One 2, Ornamentation Two 23, produces a pollen grain with a  
 228 single annulate pore, a tectate exine, spheroidal shape, monad dispersal, and  
 229 scabrate surface ornamentation. This combination represents an archetypal  
 230 grass pollen grain [19]. However, such real-world morphologies are a limited  
 231 subset of the total 3,643,200 possible combinations in this morphospace, and  
 232 consequently some of the morphologies shown in Figure 2 are theoretical and  
 233 not known to exist in nature.

234

### 235 **3.2. Patterns of morphospace occupation by real-world Neotropical** 236 **angiosperm pollen grains**

237 In order to map real-world Neotropical angiosperm pollen grains onto this  
 238 morphospace, I examined a monograph containing morphological descriptions  
 239 and images of pollen grains produced by plants growing on BCI [20]. I scored the  
 240 angiosperm pollen grains in this monograph for each of the characters used to  
 241 construct Figure 2. I excluded the following taxa from this analysis: (1) those

lacking a plate, (2) those with missing characters, and (3) polyads. I followed the higher taxonomic classifications in [20], and this resulted in a dataset consisting of 115 angiosperm plant families, which contain a total of 1,008 species and produce a total of 468 pollen morphotypes (Table 1; a full list of taxa and character scorings is provided in the Supplementary Material). This is just 0.01% of the 3,643,200 pollen morphotypes that were generated algorithmically by  $n$  choose  $k$  combination and used to construct the raw morphospace shown in Figure 2.

I split these 115 families into monocots (monocotyledons; angiosperms whose seeds contain a single cotyledon) and dicots (dicotyledons; angiosperms whose seeds contain two cotyledons). This partitioned dataset consists of 19 monocot families, which contain a total of 184 species and produce 68 pollen morphotypes, and 96 dicot families, which contain a total of 824 species and produce 407 pollen morphotypes (Table 1). I then mapped each BCI monocot and dicot pollen morphotype onto the raw morphospace shown in Figure 2. This shows how the morphospace occupied by the pollen of these two plant groups compares to the morphospace created by 3,643,200 algorithmically generated pollen morphotypes (Fig. 3).

The number of real-world BCI morphotypes in this analysis is clearly too few to occupy the entire raw morphospace. However, visual inspection of Figure 3 highlights that despite this, a considerable amount of the available morphospace is occupied by the BCI pollen morphotypes (Fig. 3). This is particularly noticeable in the morphospace occupied by dicot pollen, and for this plant group there are some associations between character states that are completely saturated. In this context, saturation may be defined as the

267 proportion of possible associations between different character states that are  
268 realized in nature. For example, there is an association between the tectate exine  
269 character state and each of the shape character states, and there is also an  
270 association between the monad dispersal state and each of the shape character  
271 states (Fig. 3*b*). Both of these examples therefore have a saturation score of 1.  
272 There is also a strong, but not saturated, association between these character-  
273 states for monocot pollen (Fig. 3*a*). For example, there are 10 possible  
274 associations between the tectate exine character state and the shape character  
275 states and six of these are realised in the BCI dataset of monocot pollen (Fig. 3*a*).  
276 This gives a saturation score of 0.6.

277 Both monocots and dicots produce pollen with a variety of surface  
278 ornamentation patterns, and for both plant groups this area of the morphospace  
279 is well explored (see the Ornamentation One axis of Fig. 3). For both of these  
280 plant groups, there is a strong association between the Ornamentation One  
281 character states and the uppermost extremity of the Ornamentation Two axis  
282 (Fig. 3). This association represents pollen grains that have just a single type of  
283 surface ornamentation. Associations between character states on the  
284 Ornamentation One axis and other character states on the Ornamentation Two  
285 axis represent pollen grains that have both primary and secondary surface  
286 ornamentation. Both monocots and dicots have evolved the means to produce  
287 pollen with both primary and secondary ornamentation, and this morphospace  
288 shows how this capability has expanded their morphological variety (Fig. 3).

289 This mapping also shows that the pollen grains produced by BCI  
290 monocots and dicots occupy strikingly different regions of morphospace (Fig. 3).  
291 This is highlighted by the differences in monocot and dicot aperture character

states. Apertures are a key character in the study of angiosperm pollen, and are thought to be a key difference between the pollen of the monocot and dicot clades [21]. The aperture character states of BCI monocots are situated in two groups that are positioned at opposite poles of the aperture character axis (Fig. 3*a*). The group at the lower extremity of the character axis contains inaperturate, monoporate, triporate, pantoporate and stephanoporate morphotypes, while the group at the upper extremity contains zonorate, trichotomosulcate, disulcate and monosulcate morphotypes (see [13] for terminology). These two groups are separated by a region of morphospace that is largely unoccupied by monocot morphotypes (Fig. 3*a*). This region contains tricolpate, stephanocolpate and tricolporate forms (see [13] for terminology), and is occupied by a variety of dicot pollen morphotypes (Fig. 3*b*). These aperture types are highly characteristic of the eudicot plant clade, but the selective pressure (if any) driving their adoption by this plant group is largely unknown [22].

## **4. Discussion**

### **4.1. Pollen grains and Neotropical plant diversity**

There are fewer pollen morphotypes than plant species in the dataset of BCI angiosperms that I have used in this paper. This is reflected in the number of pollen morphotypes expressed as a proportion of species for all angiosperms in the dataset as well as monocots and dicots (Table 1). Some of the plant families that constitute the BCI flora produce very few pollen morphotypes. Such families include the Poaceae (grasses), which are taxonomically very rich (53 species in this BCI dataset) but only produce six pollen morphotypes (Table 1). Other

317 families are also very speciose but instead produce a large number of pollen  
318 morphotypes, and an example is the Papilionoideae (a sub-family of legumes),  
319 which contains 60 species in this BCI dataset and produces 47 pollen  
320 morphotypes (Table 1).

321       This pattern, in which a plant group is characterised by the production of  
322 fewer pollen morphotypes than species, has been frequently observed in many  
323 geographic regions and time periods, and pollen grains are said to suffer from  
324 low taxonomic resolution [23]. This pattern holds for the BCI flora as a whole,  
325 and for certain families, (Table 1), but when plant species richness is plotted  
326 against the number of pollen morphotypes for all families at BCI there is  
327 considerable scatter in the distribution of data points (Fig. 4). This shows that  
328 the relationship between the taxonomic diversity of the BCI flora and the  
329 number of pollen morphotypes produced by its constituent plants is too variable  
330 to be encompassed by a single notion of low taxonomic resolution.

331       Some data points are indeed situated well below the line of equality, and  
332 pollen morphotypes clearly underestimate the species diversity of these families  
333 (Fig. 4). However, there are 32 families that plot directly on the line of equality,  
334 23 families that miss the line by just 1 (both excluding families represented by a  
335 single species on BCI), and there are several other families that plot close to this  
336 line of equality (Fig. 4). For these families, assemblages of dispersed pollen  
337 grains will provide an accurate or at least reasonable reflection of their species  
338 diversity (Fig. 4). Additionally, there are some plant species on BCI that produce  
339 dimorphic pollen, and examples include *Cocos nucifera* (the coconut tree), which  
340 produces pollen with either a monosulcate or a trichotomosulcate aperture.  
341 Indeed, among monocots there are several families in which pollen aperture type

342 is not consistent within anthers, and these include the Agavaceae and Iridaceae  
 343 (both Asparagales), as well as the Arecaceae, which contains *Cocos nucifera* [21].  
 344 It is possible, therefore, that in certain circumstances data points could lie above  
 345 the line of equality in Figure 4, and that assemblages of dispersed pollen grains  
 346 could overestimate the diversity of the source vegetation. The patterns of  
 347 morphospace occupation by BCI angiosperm pollen grains (Fig. 3) therefore  
 348 reflect a situation in which some plant families, such as the Papilionoideae,  
 349 contribute a large quantity of morphological variety, whereas other taxa such as  
 350 the Poaceae contribute much less morphological variety.

351

#### 352 **4.2. The structure and utility of a simplified palynological taxonomy**

353 In this paper, the morphology of angiosperm pollen grains has been encoded in a  
 354 system of discrete characters (see Fig. 2 and the Supplementary Material). This  
 355 system contains fewer characters than would be employed by a palynologist  
 356 seeking to classify pollen grains for the purpose of reconstruction vegetation  
 357 history, and omits certain characters that may be phylogenetically informative  
 358 such as the composition of the endexine and the nature of the intine [21].  
 359 Additionally, when observed using a microscope, the surface ornamentation of  
 360 pollen grains can appear to vary continuously between individuals and species in  
 361 a manner that is not well captured by the discrete approach that is taken here. As  
 362 such, both the total number of character combinations that are represented by  
 363 the raw morphospace shown in Figure 2, and the number of BCI pollen  
 364 morphotypes mapped onto this morphospace in Figure 3, represent minimum  
 365 diversity. The inclusion of just two size classes  $\binom{2}{1}$ , for example, would double



the total number of possible character state combinations used to construct the raw morphospace shown in Figure 2:

$$\binom{60}{1}\binom{4}{1}\binom{10}{1}\binom{6}{1}\binom{23}{2}\binom{2}{1} = 7,286,400.$$

In contrast, the removal of secondary ornamentation would substantially reduce the total number of possible character state combinations:

$$\binom{60}{1}\binom{4}{1}\binom{10}{1}\binom{6}{1}\binom{22}{1} = 316,800.$$

Secondary ornamentation is a rare character in the BCI flora (there are just 65 BCI pollen morphotypes with secondary ornamentation), and in this respect BCI is typical of angiosperm pollen morphology in other floras.

This discrete character-based system of encoding pollen morphology is comparable to formal taxonomic descriptions of pollen grains in the sense that both involve breaking pollen morphology down into individual characters, which are then described using specialist terminology (Fig. 1*a*) [13]. Further, in spite of its simplicity, this system contains a relationship between the number of species per pollen morphotype and the number of pollen morphotypes that is essentially of the same mathematical form as the relationship between the number of species per genus and the number of genera in other taxonomic systems (Fig. 5). These include classifications of plants, sponges, molluscs, fishes, beetles and birds [24, 25]. The correspondence between Figure 5 and similar plots from other taxonomic systems [24, 25], which are often presented on log-log plots, suggests that studying the structure of palynological taxonomy through evolutionary time could provide data on the diversification history of plant clades [see 26].

387           Together with data comparing plant species richness and the number of  
388 pollen morphotypes for all families at BCI (Fig. 4), Figure 5 indicates that  
389 angiosperm pollen morphospace at BCI (Fig. 3) has been filled mostly by pollen  
390 morphotypes that are unique to single species (301 instances; top-left data point  
391 in Fig. 5). There are 167 pollen morphotypes that are produced either by  
392 different species within a single family, such as the archetypal grass pollen grain,  
393 or by taxa within two or more families, and these repetitions represent 36% of  
394 the pollen morphotypes at BCI (Figure S1). This emphasizes that when studying  
395 assemblages of Neotropical angiosperm pollen grains, even a highly simplified  
396 system consisting of the five discrete characters used in this paper can capture  
397 reasonable data on the diversity of the source vegetation. The addition of further  
398 characters would likely increase the number of pollen morphotypes and move  
399 the data points in Figure 5 upwards and to the left. If these results are applicable  
400 to other regions of the world and through geological time then they provide a  
401 degree of support for studies that use pollen grains to reconstruct the diversity  
402 and composition of tropical vegetation through time.

403

#### 404   **4.3. The production and distribution of biological form**

405   In this paper I have not generated morphologies using a growth model that  
406 incorporates morphogenetic processes [cf. 6,27,28]. Instead, I have represented  
407 morphogenesis in angiosperm pollen by the algorithmic combination and  
408 enumeration of discrete morphological characters (e.g. Fig. 1), and this  
409 combinatorial approach allows for the explicit comparison of occupied and  
410 unoccupied morphospace. Each of the 3,643,200 morphotypes that can be  
411 produced from the characters I have used in this study (Fig. 2; Table 1;

Supplementary Material) is biologically plausible and structurally possible. There are no morphotypes with both tricolpate and monocolpate apertures, for example, and there are no morphotypes that contain a character state that is not known to exist in nature. However, the number of these plausible and possible angiosperm pollen morphotypes is far greater than the number of angiosperm species on Earth as a whole (~260,000; [1]) and in the BCI dataset I have used here (1008; Table 1). This means that when real-world forms are mapped onto the raw morphospace that contains these 3,643,200 morphotypes there are large areas of unoccupied morphospace (Fig. 3). Although the BCI dataset used in this study represents just a small fraction of angiosperm life, I suspect that a similar analysis incorporating all known angiosperm species would produce a similar pattern.

One way of interpreting patterns of morphospace occupation is to think of the entire morphospace as a fitness landscape, with occupied areas representing local optima and unoccupied areas representing less fit alternatives [12]. This view of morphospace occupation is informed by Wright's studies of allele combinations and fitness landscapes, in which he viewed a possible  $10^{1000}$  allele combinations as a rugged field with "an enormous number of widely separated harmonious combinations" [29, p. 358]. Wright viewed these harmonious combinations as adaptive peaks separated by maladaptive valleys, and Stebbins, without directly referencing Wright's work, seems to interpret the patterns in his floral morphospace in a similar fashion: "there are three types of combinations which, although structurally possible, are found so rarely that they are apparently poor adaptations, usually of low survival value" [3, p. 314].

436 Many attempts have been made to establish the function of the  
437 morphological characters that together make up a pollen grain [e.g. 30], and  
438 discussion has ranged from the role of surface ornamentation in pollination [22],  
439 to the role of apertures in the desiccation, hydration and germination of pollen  
440 grains [31]. Consequently, it is currently unclear which morphological features of  
441 pollen grains are adaptive and which are not, and this means that the relative  
442 roles of chance versus necessity in generating the morphological diversity of  
443 pollen grains are largely unknown. Indeed, the wide range of plausible functional  
444 roles for the morphological features of pollen grains, suggests that they are "not  
445 'optimally' designed for a specific function, but merely structures that work with  
446 varying efficiency in a specific ecological and evolutionary context" [30, p. 182].  
447 With this in mind, I do not interpret the areas of unoccupied morphospace in  
448 Figure 3 as regions containing pollen morphotypes that are maladaptive in the  
449 context of the BCI flora. It is possible, however, that further experimental work  
450 on the function of specific morphological features may reveal regions of  
451 morphospace that are characterised by forms with specific functional properties.  
452 Such work could also include analyses of the tendency of certain morphological  
453 features to arise by self-assembly processes, which may produce forms that  
454 reflect convergence upon a minimum free energy state rather than the effects of  
455 natural selection [32].

456 Alternative interpretations of morphospace occupation include the  
457 suggestion that the present-day distribution of morphology (e.g. Fig. 3) reflects  
458 an earlier more continuous distribution that has been subsequently fragmented  
459 by extinction [12], as well as Gould's classic idea that chance played a major role  
460 in distributing form, with the contingent elimination of forms early in their

461 evolutionary history producing a clumpy distribution of morphology among  
462 living organisms [12,33]. The forms of such living organisms may represent  
463 relatively small historically defined groups that are drawn from more-or-less  
464 unlimited possibilities [e.g. 3], or could instead be fundamentally limited by  
465 structural and/or phylogenetic constraints (see [5] for angiosperm pollen and  
466 [34] for animal skeletons). Results of this analysis suggest that areas in  
467 morphospace that are unoccupied by modern organisms (e.g. Fig. 3) could reflect  
468 the fact that many potentially viable morphologies have not yet been produced  
469 by the evolutionary process [35].

470       However, if a morphospace is defined using character states that are  
471 sufficiently broad in scope [e.g. 34] then the possibility of unrealized  
472 morphologies can be effectively eliminated, whereas if large numbers of  
473 character states are used to define a morphospace (as in Fig. 2), then the number  
474 of unrealized morphologies can be increased at will. Despite this scaling issue,  
475 each of the character states used to construct the raw morphospace shown in  
476 Figure 2 is known to exist in nature and is very general compared to the  
477 complexity of pollen morphology as viewed through a microscope [13].

478 Additionally, none of the 3,643,200 pollen morphotypes that these general and  
479 simplified character states produce seems biologically implausible. In this light,  
480 the possibility of large numbers of unrealized angiosperm pollen morphologies  
481 seems real, and the number of such potential forms that have not yet evolved  
482 appears to be vast. This is in contrast to analyses of animal skeletons, for which  
483 "the number of potential skeletal designs, defined in very general terms, is not  
484 inordinately large" [34, p. 350]. However, the difference in the outcomes of these  
485 two analyses (angiosperm pollen versus animal skeletons) is primarily a

486 consequence of the different scope of the characters that are defined in each  
487 case.

488         In the dataset of BCI angiosperm pollen I have analysed in this paper  
489 there are 167 morphotypes that are repeated among higher taxa (Fig S1, Fig. 6),  
490 and these repetitions highlight the numbers and kinds of morphotypes that  
491 reflect phylogenetic constraint as well as evolutionary convergence. The greatest  
492 number of these repetitions occurs among different species or genera within  
493 single families ( $n=68$ ; Fig. 6). Examples of this include the pollen of grasses  
494 (single annulate pore, tectate exine, spheroidal shape, monad dispersal, scabrate  
495 surface ornamentation; 31 repetitions in one family), laurels (inaperturate,  
496 intectate exine, spheroidal shape, monad dispersal, echinate surface  
497 ornamentation; 11 repetitions in one family) and figs (diporate, tectate exine,  
498 oblate shape, monad dispersal, psilate surface ornamentation; 7 repetitions in  
499 one family). These morphologies exhibit strong phylogenetic constraint because  
500 they are common within these families but are restricted to them.

501         In contrast, there are relatively few morphotypes that are repeated in  
502 three–nine families (Fig. 6). These 35 morphotypes are notably similar in terms  
503 of their aperture character state: 23 of these morphotypes share the same  
504 aperture character state (tricolporate; 65.7%), and all the morphotypes that are  
505 repeated among six–nine families share this character state (Fig. 6). With respect  
506 to aperture character states, this highlights a degree of nestedness in the  
507 distribution of angiosperm pollen morphology at BCI. Seven different  
508 morphotypes are repeated in both the monocot and dicot clades (Fig. 6). These  
509 morphotypes also have a very uniform set of aperture characters: two are  
510 inaperturate and five have multiple pores (pantoporate) (Fig. 6). As quoted in

511 [22] "The systematic distribution of periporate [pantoporate of many authors]  
512 pollen in extant angiosperms indicates that this aperture configuration evolved  
513 independently in many different groups" [36, p. 406]. The repetition of these  
514 aperture configurations across a fundamental divide in plant phylogeny is an  
515 example of evolutionary convergence, and in this respect, the BCI flora reflects a  
516 broader pattern in angiosperm evolution.

517

## 518 **5. Concluding remarks**

519 The combinatorial approach to angiosperm pollen morphology that I have  
520 outlined in this paper provides an example of how the general challenge of  
521 analysing large numbers of discrete character combinations [3] can be overcome.  
522 This approach generates biological forms by algorithmically combining character  
523 states and enumerating the results of these combinations (Fig. 1). I have used  
524 this approach to generate 3,643,200 angiosperm pollen morphotypes and  
525 produce a raw morphospace (Fig. 2). Some of these morphotypes correspond to  
526 real-world forms whereas others do not.

527 I have mapped the pollen of 1008 species of Neotropical angiosperms  
528 growing on BCI, Panama, onto this raw morphospace (Fig. 3). This allows  
529 realized and unrealized forms to be compared, and shows that despite the  
530 relatively small size of the BCI dataset, these real-world pollen grains occupy a  
531 considerable amount of the raw morphospace (Fig. 3). This highlights that in  
532 addition to their well-documented taxonomic diversity, Neotropical rainforests  
533 also represent an enormous reservoir of morphological diversity. Some plant  
534 families, such as the Papilionoideae, contribute a large amount of morphological

535 diversity, whereas other taxa such as the Poaceae contribute much less  
536 morphological diversity (Fig. 4).

537         In this paper, the morphology of angiosperm pollen grains has been  
538 encoded in a highly simplified system of five discrete characters. Despite this  
539 simplicity, this system contains a relationship between the number of species per  
540 pollen morphotype and the number of pollen morphotypes that has essentially  
541 the same mathematical form as the relationship between the number of species  
542 per genus and the number of genera in other taxonomic systems (Fig. 5). This  
543 highlights that angiosperm pollen morphospace at BCI (Fig. 3) has been filled  
544 mostly by pollen morphotypes that are unique to single species (Fig. 5).  
545 Angiosperm pollen morphotypes that are repeated among higher taxa at BCI  
546 reflect both phylogenetic constraint and evolutionary convergence (Fig. 6).

547

548 **Data accessibility.** The datasets and code supporting this article have been  
549 uploaded as part of the Supplementary Material.

550

551 **Competing interests.** I declare I have no competing interests.

552

553 **Author contributions.** LM conceived, designed and undertook the study. LM  
554 wrote the paper.

555

556 **Acknowledgements.** I am grateful to Claire Belcher and Carlos Jaramillo who  
557 provided useful comments on an earlier version of this work. I am grateful to the  
558 reviewers of this work, particularly Roger Thomas, whose thoughtful  
559 suggestions clarified and improved the ideas presented here.



560

561 **Funding.** I received no funding for this work.

562

563 **References**

- 564 1. Soltis PS, Soltis DE. 2004 The origin and diversification of angiosperms.  
565 *Am. J. Bot.* **91**, 1614–1626.
- 566 2. Oyston JW, Hughes M, Gerber S, Wills MA. 2016 Why should we  
567 investigate the morphological disparity of plant clades? *Ann. Bot.* **117**,  
568 859–879.
- 569 3. Stebbins GL. 1951 Natural selection and the differentiation of angiosperm  
570 families. *Evolution* **5**, 299–324.
- 571 4. Chartier M, Jabbour F, Gerber S, *et al.* 2014 The floral morphospace – a  
572 modern comparative approach to study angiosperm evolution. *New Phyt.*  
573 **204**, 841–853.
- 574 5. Lupia R. 1999 Discordant morphological disparity and taxonomic  
575 diversity during the Cretaceous angiosperm radiation: North American  
576 pollen record. *Paleobiology* **25**, 1–28.
- 577 6. McGhee GR. 1999 *Theoretical Morphology: the Concept and its*  
578 *Applications*. New York, NY: Columbia University Press.
- 579 7. Eble G. 2000 Theoretical morphology: state of the art. *Paleobiology* **26**,  
580 520–528.
- 581 8. Niklas KJ. 1999 Evolutionary walks through a land plant morphospace. *J.*  
582 *Exp. Bot.* **50**, 39–52.
- 583 9. Raup DM, Michelson A. 1965 Theoretical morphology of the coiled shell.  
584 *Science* **147**, 1294–1295.

- 585 10. Foote M. 1992 Paleozoic record of morphological diversity in blastozoan  
586 echinoderms. *Proc. Natl Acad. Sci. USA* **89**, 7325–7329.
- 587 11. Foote M. 1994 Morphological disparity in Ordovician–Devonian crinoids  
588 and the early saturation of morphological space. *Paleobiology* **20**, 320–  
589 344.
- 590 12. Erwin DH. 2007 Disparity: morphological pattern and developmental  
591 context. *Palaeontology* **50**, 57–73.
- 592 13. Punt W, Hoen PP, Blackmore S, Nilsson, Le Thomas A. 2007 Glossary of  
593 pollen and spore terminology. *Rev. Palaeobot. Palynol.* **143**, 1–81.
- 594 14. Harris JM, Hirst JL, Mossinghoff MJ. 2000. *Combinatorics and Graph*  
595 *Theory*. New York, NY: Springer-Verlag.
- 596 15. Glatz E, Mavromatidis S, Ager B, Dimitropoulos X. 2012 Visualizing Big  
597 Network Traffic Data using Frequent Pattern Mining and Hypergraphs.  
598 *Computing* **96**, 27–38.
- 599 16. Tricaud S, Nance K, Saadé P. 2011 Visualizing network activity using  
600 parallel coordinates. *Proc. Hawaii International Conference on System*  
601 *Sciences* **44**, 1–8.
- 602 17. ter Steege H, Pitman NCA, Sabatier D, *et al.* 2013 Hyperdominance in the  
603 Amazonian tree flora. *Science* **342**, 1243092.
- 604 18. Hubbell SP, Foster RB, O'Brien ST, Harms KE, Condit R, Wechsler SJ,  
605 Wright S, Loo de Lao S. 1999 Light-gap disturbances, recruitment  
606 limitaiton, and tree diversity in a Neotropical forest. *Science* **283**, 554–  
607 557.

- 608 19. Mander L, Li M, Mio W, Fowlkes CC, Punyasena SW. 2013 Classification of  
609 grass pollen through the quantitative analysis of surface ornamentation  
610 and texture. *Proc. R. Soc B* **280**, 20131905.
- 611 20. Roubik DW, Moreno JE. 1991 *Pollen and spores of Barro Colorado Island*.  
612 St. Louis, Missouri: Missouri Botanical Gardens.
- 613 21. Harley MM, Zavada MS. 2000 Pollen of the monocotyledons: selecting  
614 characters for cladistic analysis. In *Monocots: Systematics and Evolution*  
615 (eds KL Wilson, DA Morrison), pp. 194–213. Melbourne, Australia: CSIRO.
- 616 22. Chaloner WG. 2013 Three palynological puzzles. *Int. J. Plant Sci.* **174**, 602–  
617 607.
- 618 23. Mander L, Punyasena SW. 2014 On the taxonomic resolution of pollen and  
619 spore records of Earth's vegetation. *Int. J. Plant Sci.* **175**, 931–945.
- 620 24. Willis JC, Yule GU. 1922 Some statistics of evolution and geographical  
621 distribution in plants and animals and their significance. *Nature* **109**,  
622 117–179.
- 623 25. Burlando B. 1990 The fractal dimension of taxonomic systems. *J. Theor.*  
624 *Biol.* **146**, 99–114.
- 625 26. Foote M. 2012 Evolutionary dynamics of taxonomic structure. *Biol. Lett.* **8**,  
626 135–138.
- 627 27. Thompson DW. 1942 *On Growth and Form*. 2nd Edition. London, UK:  
628 Cambridge University Press.
- 629 28. Turing AM. 1952 The chemical basis of morphogenesis. *Phil. Trans. R. Soc.*  
630 *Lond. B* **237**, 37–72.
- 631 29. Wright S. 1932 The roles of mutation, inbreeding, crossbreeding, and  
632 selection in evolution. *Int. Congr. Genet.* **6** **1**, 356–366.

- 633 30. Blackmore S, Ferguson IK. eds 1986 *Pollen and Spores: Form and Function*.  
 634 London, UK: Academic Press.
- 635 31. Katifori E, Alben S, Cerda E, Nelson DR, Dumais J. 2010 Foldable  
 636 structures and the natural design of pollen grains. *Proc. Natl Acad. Sci. USA*  
 637 **107**, 7635–7639.
- 638 32. Hemsley AR, Collinson ME, Kovach WL, Vincent B, Williams T. 1994 The  
 639 role of self-assembly in biological systems: evidence from iridescent  
 640 colloidal sporopollenin in *Selaginella* megaspore walls. *Phil. Trans. R. Soc.*  
 641 *Lond. B* **345**, 163–173.
- 642 33. Gould SJ. 1989 *Wonderful Life*. New York, NY: Norton.
- 643 34. Thomas RDK, Reif W-E. 1993 The skeleton space: a finite set of organic  
 644 designs. *Evolution* **47**, 341–360.
- 645 35. Lewontin RC. 2003 Four complications in understanding the evolutionary  
 646 process. *Santa Fe Institute Bulletin* **18**, (centre insert, unpaginated).
- 647 36. Friis EM, Crane PR, Pedersen KR. 2011 *Early Flowers and Angiosperm*  
 648 *Evolution*. New York, NY: Cambridge University Press.

649

## 650 **Figure and table captions**

651 **Figure 1.** (a) Schematic illustration of characters that make up an angiosperm  
 652 pollen grain. Aperture character states: inaperturate (1), triporate (2), tricolpate  
 653 (3). Exine character states: atectate (1), tectate (2). Shape character states:  
 654 oblate (1), spheroidal (2), prolate (3). Dispersal character states: monad (1),  
 655 tetrad (2). Surface ornamentation character states: baculate (1), echinate (2),  
 656 clavate (3), reticulate (4), no ornamentation (5). Different pollen morphotypes  
 657 reflect different combinations of these character states. (b) Binomial coefficients

658 for each character. (c) Showing how these different combinations can be  
659 represented using a parallel coordinates plot.

660

661 **Figure 2.** Parallel coordinates plot of 3,643,200 angiosperm pollen morphotypes  
662 generated algorithmically by  $n$  choose  $k$  combination. Numbers enclosed within  
663 white circles represent each of the character states used to produce this  
664 morphospace. Only the first and last aperture and surface ornamentation  
665 character states are labelled to prevent overcrowding of these axes. A full list of  
666 character states corresponding to these numbers is provided in the  
667 Supplementary Material.

668

669 **Figure 3.** Parallel coordinates plots showing how pollen grains produced by  
670 angiosperms growing on Barro Colorado Island map onto the morphospace  
671 shown in Figure 2. In both plots this morphospace is shown in light grey, and is  
672 overlain by (a) monocot pollen morphotypes in dark blue and (b) dicot pollen  
673 morphotypes in orange. A full list of taxa and character scorings is provided in  
674 the Supplementary Material.

675

676 **Figure 4.** Scatterplot comparing the species richness of angiosperm families  
677 from Barro Colorado Island with the number of pollen morphotypes produced by  
678 each of these families. Datapoints have been plotted for each of the 19 monocot  
679 and 96 dicot families analysed in this paper but datapoints for certain families  
680 overlap. Both axes are plotted on a logarithmic scale. Diagonal dashed line  
681 represents a line of equality.

682

**Figure 5.** Scatterplot showing the hollow-curve relationship between the number of species per angiosperm pollen morphotype and the number of angiosperm pollen morphotypes on Barro Colorado Island.

**Figure 6.** Proportions of angiosperm pollen morphotypes that are repeated among higher taxa at BCI (green vertical bars). Numbers enclosed within grey vertical boxes show the distribution of aperture character states that are repeated among higher taxa at BCI (see dataset S1 for a full list of aperture character states). Key aperture character states repeated among three–nine families as follows: Inaperturate (1), Pantoporate (11), Tricolpate (25), Tricolporate (44), Tricolporate Marginate (45), Pantocolporate (48), Monosulcate (54). Cartoons to the right are schematic pollen grains that graphically display each of these aperture character states.

**Table 1.** Summary of the data and results discussed in this paper.

**Supplementary material**

**Dataset S1.** List of characters used to generate the raw morphospace shown in Figure 2.

**Dataset S2.** Taxa and character scorings of BCI pollen grains used to generate Figure 3. Refer to Dataset S1 for list of characters.

**Supplementary File S1.** An example Python algorithm that counts and enumerates all possible combinations of the character states shown in Figure 1a.

708

709 **Supplementary File S2.** An example of a .csv file containing all possible  
710 combinations of the character states (in numerical form) shown in Figure 1*a*.

711

712 **Figure S1.** Barchart of the frequency of angiosperm pollen morphotypes at BCI.

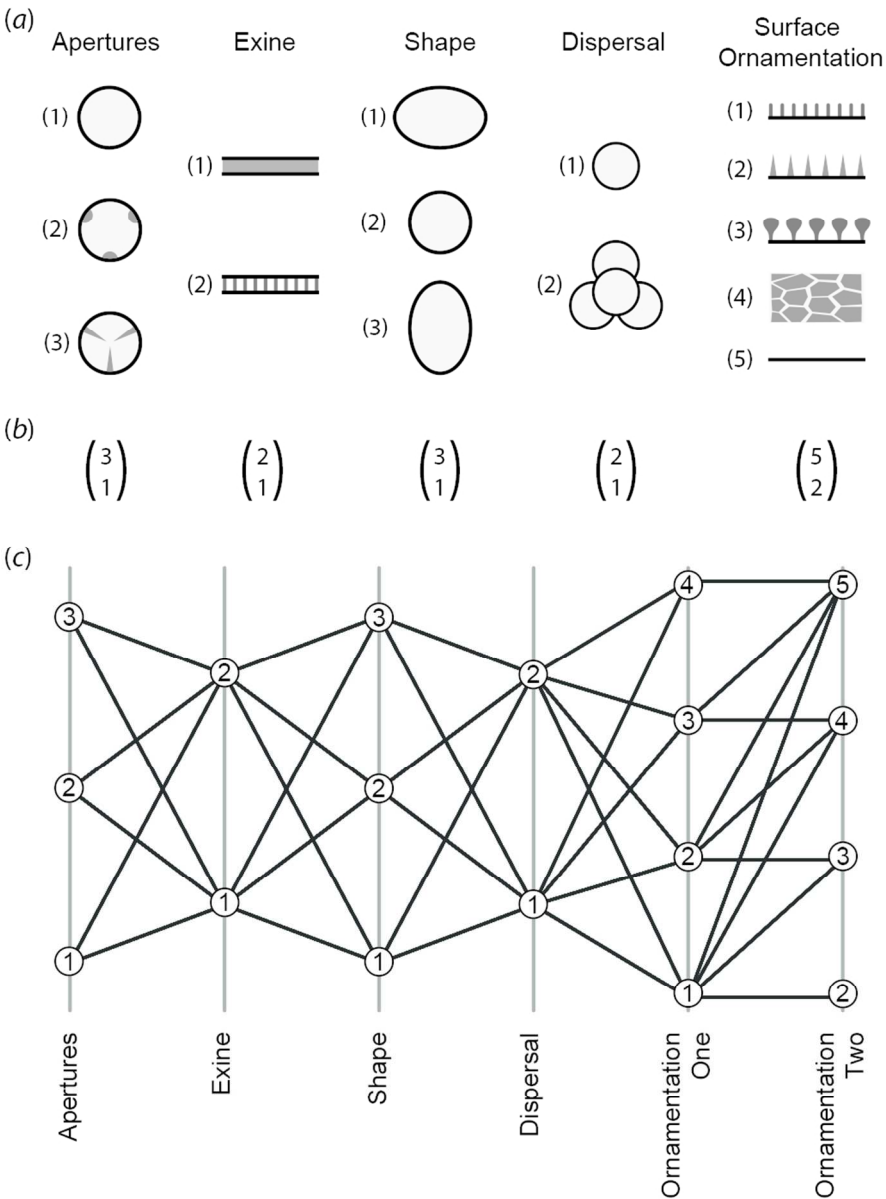


Figure 1. (a) Schematic illustration of characters that make up an angiosperm pollen grain. Aperture character states: inaperturate (1), triporate (2), tricolpate (3). Exine character states: atectate (1), tectate (2). Shape character states: oblate (1), spheroidal (2), prolate (3). Dispersal character states: monad (1), tetrad (2). Surface ornamentation character states: baculate (1), echinate (2), clavate (3), reticulate (4), no ornamentation (5). Different pollen morphotypes reflect different combinations of these character states. (b) Binomial coefficients for each character. (c) Showing how these different combinations can be represented using a parallel coordinates plot.

88x118mm (300 x 300 DPI)



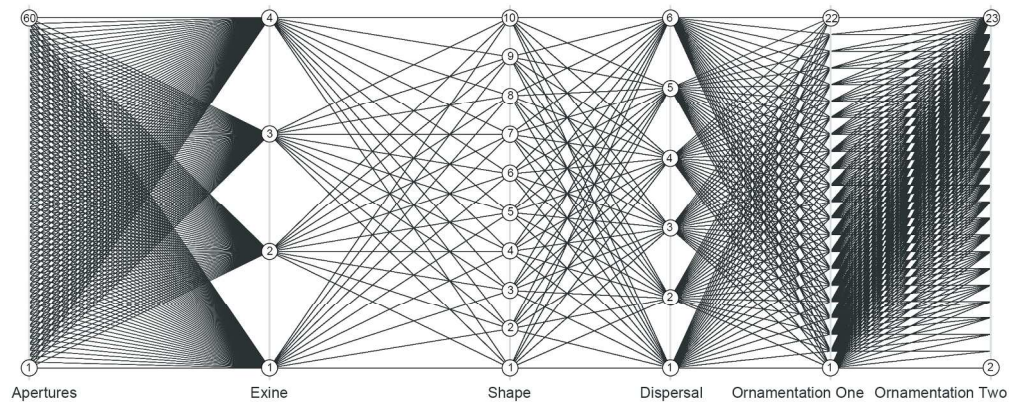


Figure 2. Parallel coordinates plot of 3,643,200 angiosperm pollen morphotypes generated algorithmically by  $n$  choose  $k$  combination. Numbers enclosed within white circles represent each of the character states used to produce this morphospace. Only the first and last aperture and surface ornamentation character states are labelled to prevent overcrowding of these axes. A full list of character states corresponding to these numbers is provided in the Supplementary Material.

187x75mm (300 x 300 DPI)

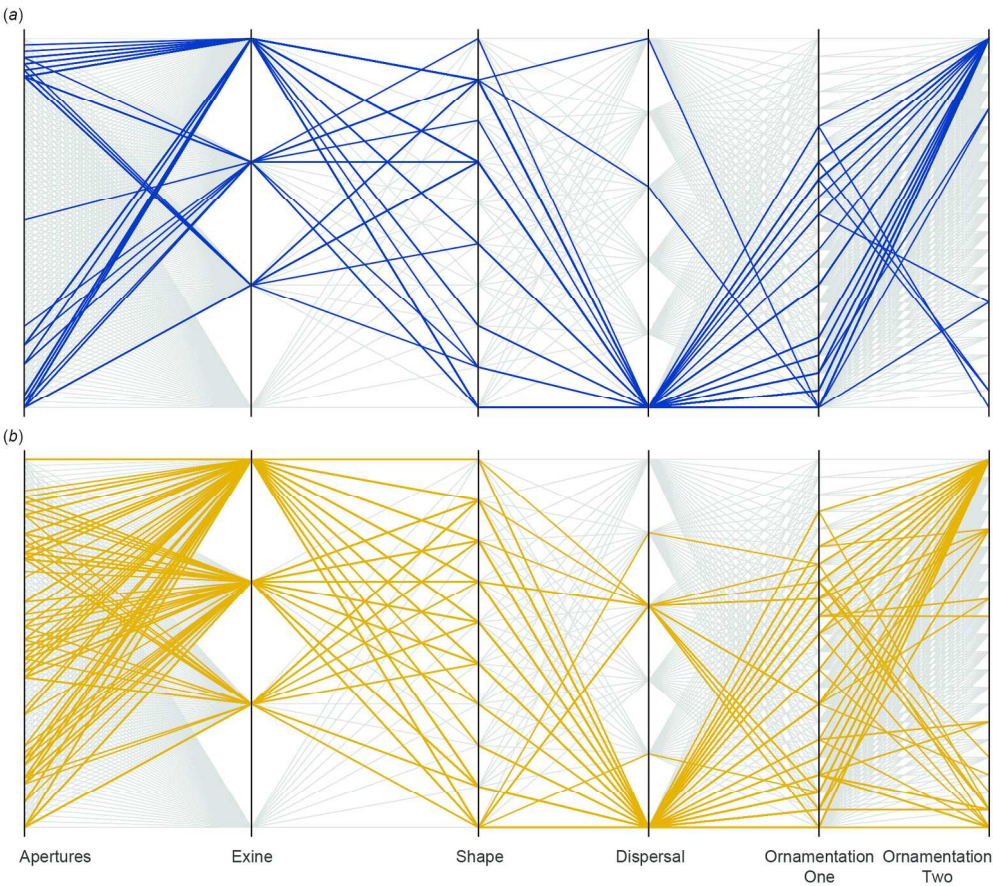


Figure 3. Parallel coordinates plots showing how pollen grains produced by angiosperms growing on Barro Colorado Island map onto the morphospace shown in Figure 2. In both plots this morphospace is shown in light grey, and is overlain by (a) monocot pollen morphotypes in dark blue and (b) dicot pollen morphotypes in orange. A full list of taxa and character scorings is provided in the Supplementary Material.

170x151mm (300 x 300 DPI)

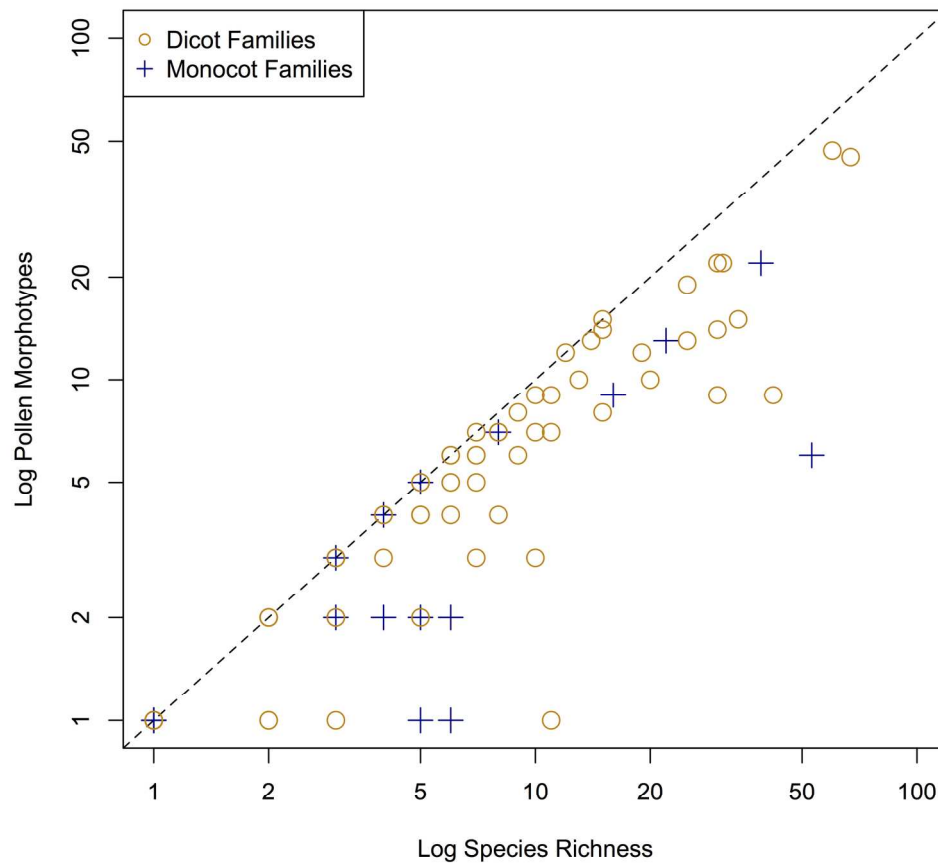


Figure 4. Scatterplot comparing the species richness of angiosperm families from Barro Colorado Island with the number of pollen morphotypes produced by each of these families. Datapoints have been plotted for each of the 19 monocot and 96 dicot families analysed in this paper but datapoints for certain families overlap. Both axes are plotted on a logarithmic scale. Diagonal dashed line represents a line of equality.

177x177mm (300 x 300 DPI)

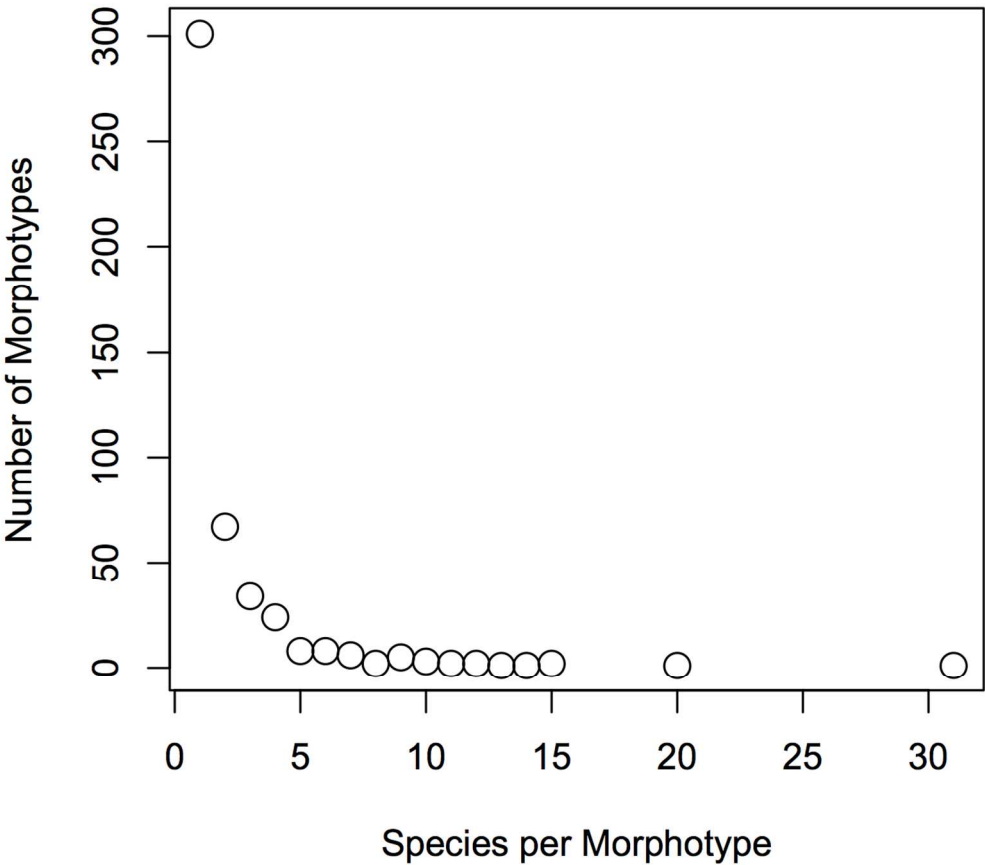


Figure 5. Scatterplot showing the hollow-curve relationship between the number of species per angiosperm pollen morphotype and the number of angiosperm pollen morphotypes on Barro Colorado Island.

117x103mm (300 x 300 DPI)

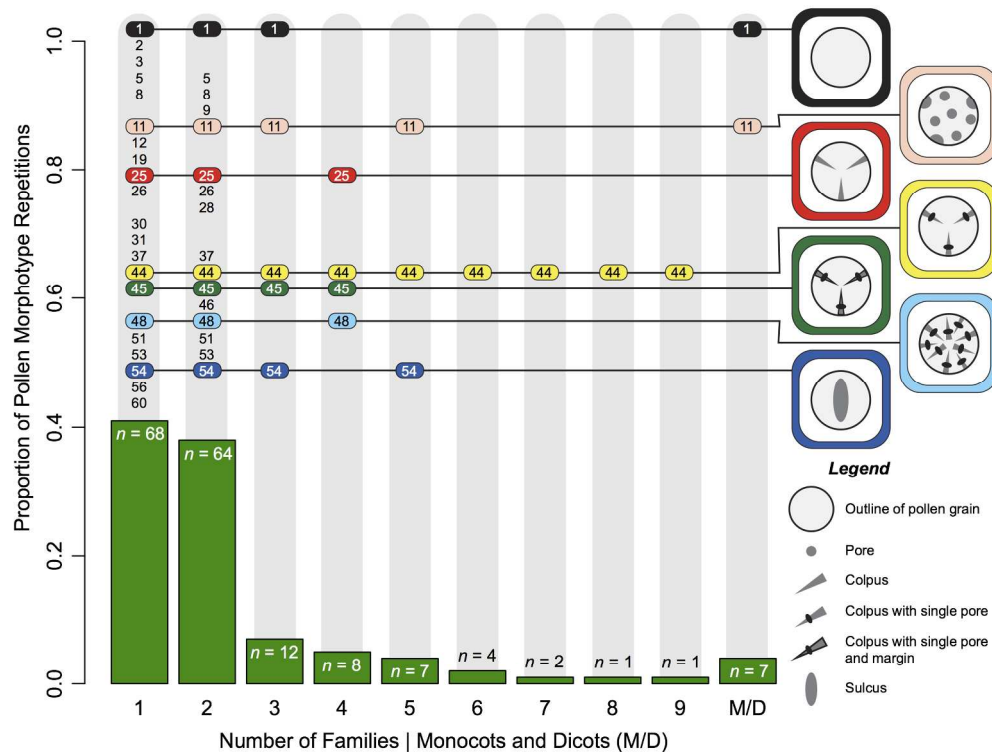


Figure 6. Proportions of angiosperm pollen morphotypes that are repeated among higher taxa at BCI (green vertical bars). Numbers enclosed within grey vertical boxes show the distribution of aperture character states that are repeated among higher taxa at BCI (see dataset S1 for a full list of aperture character states). Key aperture character states repeated among three–nine families as follows: Inaperturate (1), Pantoporate (11), Tricolpate (25), Tricolporate (44), Tricolporate Marginate (45), Pantocolporate (48), Monosulcate (54). Cartoons to the right are schematic pollen grains that graphically display each of these aperture character states.

204x153mm (300 x 300 DPI)

	Families ( <i>n</i> )	Species ( <i>n</i> )	Morphotypes ( <i>n</i> )	Morphotypes as a proportion of Species
<i>n</i> choose <i>k</i> Morphospace	-	-	3,643,200	-
BCI Angiosperm pollen	115	1008	468	0.46
BCI Monocots	19	184	68	0.37
BCI Dicots	96	824	407	0.49
BCI Araceae (monocot)	-	39	22	0.56
BCI Poaceae (monocot)	-	53	6	0.11
BCI Papilionoideae (dicot)	-	60	47	0.78
BCI Rubiaceae (dicot)	-	67	45	0.67